

Policy analysis of the English graduation benchmark in Taiwan

CHIH-MIN SHIH

Nanyang Technological University

To nudge students to study English and to improve their English proficiency, many universities in Taiwan have imposed an English graduation benchmark on their students. This article reviews this policy, using the theoretic framework for education policy analysis proposed by Haddad and Demsky (1995). The author presents relevant research findings, and concludes that this policy does not achieve the intended goals. It is suggested that policymakers reconsider this English graduation benchmark.

Keywords: English graduation benchmark, Taiwan, university students, washback, policy analysis

Introduction

In order to embrace globalisation and enhance students' proficiency in English, it has become common practice in Taiwan to impose an English graduation benchmark on university students as one of their degree requirements. This paper analyses this policy, using the theoretic framework for education policy analysis proposed by Haddad and Demsky (1995). This framework encompasses seven policy-planning processes: "(1) analysis of the existing situation; (2) the generation of policy options; (3) evaluation of policy options; (4) making the policy option; (5) planning of policy implementation; (6) policy impact assessment; (7) subsequent policy cycles" (Haddad & Demsky, 1995:24). These seven processes also form the skeleton of this article. Since there is a paucity of research on this subject in Taiwan, all relevant empirical studies are cited to present the current understanding of this issue.

Analysis of the existing situation

The economic and educational contexts played an important role in forming the policy of the English graduation benchmark. In terms of the economy, over 80% of Taiwan's gross national product depended on international trade (Taiwan Government Information Office, 2004). In 2007, Taiwan was ranked the 16th largest exporter and the 17th largest importer in the world (Taiwan Government Information Office, 2008). Therefore, foreign languages such as English, German, French, Spanish, and Japanese were used as a medium of communication with foreign companies and played an indispensable role in Taiwan's economy. In the past two decades English has become a *lingua franca* in international trade.

Starting from the mid-1980s, the higher education sector in Taiwan has witnessed an unprecedented expansion. According to the Ministry of Education (n.d.), higher education used to be under government's stringent control. In 1994, the University Law was enacted, granting universities more academic freedom and institutional autonomy. Due to the liberalism of higher education, the number of universities increased by 3.75 times, and that of students by 2.52 times from the mid-1980s to 2003. The market-oriented mechanism was also introduced to the education system (Chou, 2008). Due to a high demand for foreign-language talents, numerous universities established their English-related departments (e.g., applied foreign language department, applied English department).

The expansion of the higher education, however, has sparked unprecedented concerns about the quality of university students in the education system and society. Decades ago, being admitted to university was dubbed by the Taiwanese as a "narrow gate", which denotes the competitiveness of the Joint College Entrance Examination (JCEE). Only top-notch students were able to excel in the JCEE. In the wake of the liberation of higher education, university students' academic performance varied significantly (Shih, 2007).

The generation of policy options

Given that English has become a language universally used for international trade, and that university students' academic performance, on average, plummeted, there was a call for improving all citizens' proficiency in English. Imposing an English graduation benchmark was one of a multitude of proposals to improve proficiency in English.

Prior to the new millennium, English was a compulsory subject for university students, but the English graduation benchmark was prescribed to a limited extent at Taiwanese universities. A dramatic turn occurred after the Language Training and Testing Centre (LTTC), commissioned by the Taiwan Ministry of Education (MOE), phased in the General English Proficiency Test (GEPT¹). The introduction of the GEPT made the English graduation benchmark feasible partly due to its relatively low registration fees and five levels of proficiency (Shih, 2010); university authorities could choose one proficiency level out of five which was deemed appropriate for their students, and did not have to be concerned that students could not afford the registration fees. Some universities started to impose an English requirement on their students circa 2002. For example, since 2003, freshmen at National Chiao Tung University have been required to pass the first stage of the GEPT's high-intermediate level (Chen, 2008). This policy became contagious, and was soon adopted by many universities.

Evaluation of policy options

Numerous policies might be conducive to English learning, and they can be implemented concurrently. Therefore, I will not compare this policy with others in this article. Rather, I will focus on how universities evaluated the policy of the English graduation benchmark.

Different universities and programmes have various views of this policy. One study examined how higher education institutions (HEIs) evaluated this policy and arrived at different conclusions. Shih's (2010) research in two applied foreign language departments at different institutes of technology in 2004 examined why one department (Department A) had not prescribed an English graduation benchmark, whereas the other (Department B) had. Although faculty members in Department A believed that the graduation benchmark would motivate students to learn English, they had other concerns, namely that parents might oppose this requirement, and that students' proficiency in English was too low to pass the graduation benchmark. They also learned that this policy had not been successfully implemented at other universities, and understood that the MOE did not allow universities to prescribe this requirement in the early 2000s. Therefore, they decided not to implement this policy.

Faculty members in Department B had a different approach to this requirement. They believed that this policy would persuade students to study English, and that an English certificate would be conducive for students to seeking employment after graduation. In addition, they perceived that this requirement would make the university more competitive in the higher education market in Taiwan. School authorities could also be more confident of students' proficiency in English if they could pass the test. Although faculty members in Department B were also concerned that this requirement may flout the law of the MOE, they solved this problem by providing failing students an opportunity to take a make-up examination administered by the department. Despite its narrow scope, this research exemplified how faculty members at different universities evaluated the policy of the English graduation benchmark from various perspectives.

Validity issues

With the exception of the empirical study which demonstrated how this policy was examined at different institutions, one theoretical issue which emerges when this policy is evaluated is the validity of the test used. Since the GEPT is probably the most universally used test for graduation benchmark purposes, its validity deserves close attention. According to Messick (1996:245), "validity is an overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment". This

definition implies that interpretations of test scores are important to establish the validity of the test. One source of invalidity of the GEPT is its proficiency descriptions for different levels which fail to provide proper interpretations of test scores. Both Shih (2008b) and Vongpumivitch (2010) pointed out that this aspect is the weak link of the GEPT. For example, the LTTC (n.d.-b) claimed that a high-intermediate level of certificate roughly equates with non-English-major university graduates' proficiency in English. This statement is untenable because there is no unified English curriculum among Taiwanese universities. Therefore, the so-called proficiency in English of non-English-major university graduates is arbitrary. Due to the inaccuracy of the proficiency descriptions of the GEPT, university authorities cannot use them to correctly interpret the results of the test, and prescribe an appropriate graduation benchmark.

The other source of problematic proficiency descriptions is that the GEPT used can-do statements to portray what test takers should be able to perform in the four skills in real-world situations. Roeber and Pan (2008) pointed out that these can-do statements were not empirically substantiated and thus dubious. For example, when describing the listening tasks that test takers who pass the advanced level can perform, the LTTC (n.d.-a) states that they are able to comprehend conversations of a variety of genres, such as television programmes and debates. In their workplace, they are able to understand oral reports and negotiations when they are in a meeting or are engaged in a discussion. Such sweeping statements can be found in all of the four skills across all levels of the GEPT. However, the test tasks of the GEPT do not always reflect the claimed language uses. According to Messick (1996:244), this threat to validity is known as "construct under-representation", meaning that "the test is too narrow and fails to include important dimensions or facets of focal constructs".

Ethical issue

The second theoretical issue which deserves more attention is ethical concerns. Using a single test for gate-keeping purposes is considered inappropriate. In his article *The ethics of gate-keeping tests: What have we learned in a hundred year?*, Spolsky (1997) states that there is an inevitable psychometric error embedded in tests. Instead of eradicating it, psychometrists are acknowledging and accepting its existence, paying attention to how the results of the tests are interpreted and used, and employing several assessment methods in the process of making a critical decision. He concludes: "Tests can provide valuable data for gate-keeping decisions, but must not be left as the sole arbiter" (Spolsky, 1997:246-247). Similar suggestions have been proposed by other scholars (see Brown & Hudson, 1998; Shohamy, 2001; Brown, 2004). It appears that all of the above-mentioned literature rebuts the legitimacy of adopting a single English test as the benchmark for graduation.

Making the policy decision

As reported earlier, imposing this requirement on students was even forbidden by the Taiwan MOE in early 2000, according to the accounts of two chairpersons of applied foreign language departments at Taiwan institutes of technology in 2004 (Shih, 2010). The government's attitude towards this policy changed significantly in 2005. As part of the governmental four-year plan for 2005-2008, the Taiwan MOE mapped out the goals for improving the English ability of students at HEIs (MOE, 2005). By 2007, 50% of students at both institutes of technology and universities of technology would pass the elementary level of the GEPT, and the same percentage of university students would be certified at the intermediate level of the GEPT.

Almost concurrently, the MOE commissioned the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT), Taiwan Assessment and Evaluation Association (TWAEA) and National Yunlin University of Science and Technology to evaluate all academic programmes, a practice that would be institutionalised to hold HEIs accountable. Programmes with a negative evaluation report are to receive less funding, to recruit fewer students, and to charge less tuition as a penalty. In addition, these programmes will receive extensive media coverage, which undoubtedly will ruin their reputation. If the issues pointed out by programme reviewers are not improved, the programmes will be closed after consecutive negative evaluation results. It is obvious that programmes face dire repercussions if they fail

their evaluation. It appears that imposing an English graduation benchmark on university students has become an important index when academic programmes are reviewed. For example, it was suggested that the Department of Applied English and the Department of Accounting Information Systems of National Taichung Institute of Technology establish an English graduation benchmark (National Yunlin University of Science and Technology, 2008), an indication that this requirement should be imposed on English-major and non-English-major students. In their endeavour for acceptable evaluation reports, universities are left no choice but to lay down this requirement even if it is not endorsed by faculty members and students.

Planning of policy implementation

At national level, the MOE officials had several strategies to achieve the above-mentioned goals in the governmental four-year plan for 2005-2008 (MOE, 2005):

1. The cross-reference table of different English tests can be used as the criteria for reference.
2. Universities are encouraged to offer English programmes as students' electives, and students to take different levels of English tests.
3. Students' passing of English tests is included as one of the indices of university evaluation.
4. Universities' reward rules for improving students' foreign language proficiency should be examined and rectified, and universities should be guided to actively promote English tests and remedial teaching.
5. English competition campaigns should be held for technological institutes.
6. English teaching and learning resource centres should be established in the geographic regions of northern, southern, eastern, and central Taiwan.

At university level, close scrutiny revealed that every university has its distinct English graduation benchmark for students' proficiency in English. For example, distinguished universities (e.g., National Taiwan University, National Chiao Tung University), in general, require students to pass the first stage of the GEPT's high-intermediate level; others may set a lower criterion such as the GEPT's elementary or intermediate level. Irrespective of their benchmark, they may allow students to take other English tests (e.g., IELTS,² TOEFL,³ TOEIC⁴) as the equivalent. Upon reaching a stipulated score in one of the above-mentioned tests, students are also considered to have fulfilled this requirement. To assist and encourage students to achieve the English graduation benchmark, English self-study centres have been established, test-preparation courses have been offered, and various incentives (e.g., monetary incentives) have been awarded to students who pass the test at many HEIs. In fact, as university authorities realise that not every student can pass the graduation benchmark, they offer a "back door" (e.g., remedial English course, make-up examination) to students who fail the English test. As long as students pass the alternative, they can still graduate from the university. Such a measure is so universal that even National Taiwan University, which always recruits the best performing senior high school students in Taiwan, cannot avert. In fact, empirical studies have seldom been conducted to show that passing the course is tantamount to attaining the graduation benchmark. More often than not, these alternatives are easier options, which might be a source of invalidity.

Policy impact assessment

Stakeholders' acceptance is a critical factor which determines the success of a testing policy (Berry & Lewkoxic, 2000). It is thus necessary to investigate students' views of this requirement. It appeared that the majority of the participants in Shih's (2008a) and Chu's (2009) studies supported the policy of prescribing an English graduation benchmark. They upheld it because it motivated students to study English and thereby improved their English (Shih, 2008a; Chu, 2009); it was conducive in the context of globalisation (Chu, 2009); it reflected an inevitable global trend (Chu, 2009); it built the reputation of the university (Chu, 2009), and, English certificates were critical for future job-hunting (Shih, 2008a; Chu, 2009). However, some of the participants opposed this requirement for several reasons. First, they believed that taking an English test is a personal matter, and that universities should not prescribe any graduation benchmark (Shih, 2008a). Secondly, they disapproved of this policy because not all students would pursue

a career which required using English (Shih, 2008a; Chu, 2009). According to some participants, the requirement was an unnecessary burden for students who would not encounter foreigners in their jobs. Thirdly, some students did not consider this policy to be an appropriate way to improve English (Chu, 2009). They suggested offering more English courses and creating an English learning environment instead. Fourthly, students' proficiency in English varied; thus having the same benchmark imposed on everybody was not fair (Shih, 2008a; Chu, 2009). Fifthly, students considered the GEPT to be a test that should be used in the workplace, not in the education system (Shih, 2008a).

One relevant perspective which is worth considering is Shohamy's (2001) theory of critical language testing. She advocated that test takers be empowered to question the uses and methods of the test, and that they have the right to reject the test if they perceive that it does not fit into their knowledge. Her theory emphatically refutes the appropriateness of the English graduation benchmark and appears to promote a neutral policy which allows students to decide whether they will take the test or not. This neutral policy, in my view, is also a good solution to students' divided opinions of the controversial English graduation benchmark.

Washback studies

Probably the most direct way to analyse the impact of this policy is to conduct washback studies. To date, several research projects have focused on this subject. For example, Shih (2007) investigated the washback effects of the GEPT on students' English learning. The applied foreign language department of a university of technology (Department A) and that of an institute of technology (Department B) were selected as his research sites. When he collected his data in 2004, Department A Students did not have to take any English test whereas Department B students had to pass the first stage of the GEPT's intermediate level one year prior to graduation. Those who failed the GEPT were eligible to take the departmental make-up examination. If students failed both examinations, they would not be able to graduate. Shih (2007) found that the GEPT had limited or no impact on students' English learning in both departments, regardless of the status of the GEPT. Although the test seemed to encourage some autonomous learners when they passed the test with ease, it also caused frustration among students who failed the test. For example, English was originally the favourite subject of Super Killer, a student who participated in Shih's study. After he failed the GEPT on three successive occasions, his confidence suffered a serious setback, and his motivation was severely blunted.

Shih (2007) further deciphered why the GEPT was perceived as a high-stakes test by a marginal majority of his participants but induced limited washback. He enumerated nine factors that attenuated the washback of the GEPT on students' English learning:

1. Passing the GEPT was not desperately needed for students who would graduate in a few years' time, so they could take a lax attitude to prepare for the test.
2. Since students in his study were English majors, their routine English learning could be used as their test preparation.
3. Although students were willing to gear up for the test, they did not know how to prepare for it, or they did not have a partner with whom to practise.
4. Students were so proficient in English that they did not have to prepare for the test.
5. Students were occupied with other duties or obligations (e.g., part-time jobs) prior to the test, so they could not fully commit to their test preparation.
6. Students acknowledged that their laziness prevented them from preparing for the test.
7. Students were not autonomous learners, and did not have good learning attitudes.
8. Students abhorred testing and were reluctant to study solely for the test.
9. Department B had a make-up examination which had a loophole. In fact, the make-up examination was randomly selected from 15 sets of simulated GEPT questions to which all students had access. Diligent students could try to answer the questions and pass answer keys to those who needed to take the test to help them pass the test without any preparation.

Chu (2009) conducted another study to scrutinise the impact of the GEPT on students' English learning at two universities of technology. She elicited help from a public university of technology (University A) and a private one (University B), both of which required their day-division non-English-major students to pass the first stage of the GEPT's intermediate level. None of the 32 students interviewed told her that the GEPT had induced a high level of washback on their English learning; 19 reported that the GEPT had no impact at all. Similarly, findings which emerged from 606 copies of student questionnaires showed that the positive washback on learning was minute. Regrettably, the graduation benchmark had induced negative feelings such as fear and frustration among students. Overall, the results suggested that the English graduation benchmark had sparked more unintended, deleterious washback than the intended, positive one. Chu (2009) attributed the low washback to different factors for students with various proficiency levels. She found that the English benchmark was too formidable and daunting a task for low-achieving students to attain. The students were of the opinion that they were doomed to fail the test. Although willing to prepare for the test, they did not know how to. High-achieving students were not swayed by the English test because they studied English out of their own interests. They would study by themselves irrespective of whether there was an exit test or not. Those students whose proficiency in English was between the two extremes would wait until the last minute to prepare for the test. Their concern was to boost their test score, not their proficiency in English. The other factor that attenuated the washback on learning was that the GEPT did not meet the students' needs. For example, a participant mentioned in the interview that he was interested in improving his speaking skills, which were not assessed in most of the tests in Taiwan, including the first stage of the GEPT's intermediate level.

Chen (2008) investigated the washback of the GEPT at a leading university where undergraduate students were required to pass the first stage of the GEPT's high-intermediate level as their graduation benchmark, or they had to take remedial English courses. Results showed that students were not very motivated to study for the test; 101 participants prepared themselves for the GEPT, whereas the remaining 182 students took the test unprepared. Vongpumivitch (2006) conducted the other washback study at another prestigious university in Taiwan. She found that the GEPT was solely one of the numerous tests which students took in their lives. Regardless of the GEPT's presence, students' learning methods remained invariable. She concluded that the GEPT neither induced any long-term washback on students' English learning nor promoted life-long learning of English.

To recapitulate, the above empirical studies indicate that the GEPT induced limited or no washback on both English-major and non-English-major students, national and private HEIs, comprehensive universities, and institutions of technology. No studies, to the best of my knowledge, have shown that the English graduation benchmark generated a high level of washback on students' English learning, and therefore boosted students' proficiency in English at Taiwan HEIs. Although all washback research projects are case studies which lack generalisability, they showed limited washback effects of this policy to some extent.

Impact on the workplaces

Tests are now considered social practice (see McNamara & Roever, 2006; McNamara, 2008). Another method to examine the impact of the policy is to investigate whether an English certificate is indispensable for students' future job applications and their workplaces. Two recent studies examined this topic. Pan (2009) recruited 19 participants responsible for hiring employees in their companies whose workforces ranged from 8,000 to 30,000. She noted that only 13% of the participants required their job applicants to submit their English certificates despite their unanimous espousal of the English graduation benchmark. She also noted that 53% of her participants considered applicants who possessed an English certificate diligent. This attribute, however, was not assessed by English tests, a fact which meant that these stakeholders had a different view of the test than did testers. Her study implied that English might not be a cardinal skill when university graduates seek a job. In addition, the results of the English tests might be misinterpreted and therefore be used for construct-irrelevant purposes.

Chu (2009) focused solely on jobs that required employees to use English, and interviewed six representatives from the industry to explore students' future needs in their workplaces. Notwithstanding the important roles English might play in the workplace, professional knowledge, not English proficiency, was the determining factor when job applicants were considered for recruitment. English skills became essential when the rest of the qualifications were roughly equivalent. In addition, Chu (2009) noted that verbal and written communications with foreign clients as well as datasheet reading were the primary tasks that students would have to perform in their future workplaces. This finding was congruent with the student participants' perceptions of their future needs for English. To complete the above tasks, fundamental listening, speaking, reading, and writing skills in English as well as specialised English vocabulary knowledge were required. Chu further compared students' English needs in the workplace with the content of the GEPT and found that there was a mismatch between them. She claimed that this disparity contributed to the GEPT's lack of washback on students' learning.

Although demonstrating that English might be important in certain companies, Chu (2009) indicated that English was not the primary consideration when companies recruited new employees. If conclusions are to be drawn from these empirical studies, other core abilities which Taiwanese companies regarded highly, rather than the English graduation benchmark, should be prescribed. However, both qualitative studies involved a small number of participants. More large-scale research is required to investigate this issue.

Subsequent policy cycles

According to Haddad and Demsky (1995), the above processes of proposing, deciding, implementing, and evaluating a policy should be iterative. At national level, it appears that this policy, to date, has not been revisited and that educational authorities have not questioned its efficacy. At school level, university authorities and faculty members have attempted to fine-tune this policy. The two departments in Shih's (2010) study have undergone several modifications in their policy of English graduation benchmark, an indication that policymaking is a complex issue. For example, in 2003, faculty members in Department A decided to require all new students to pass the intermediate level of the GEPT prior to graduation. However, due to various considerations discussed earlier, this policy was not implemented. To date, a dearth of research has investigated the evolution of this policy; more studies can be conducted to further examine this issue.

Conclusion

It is not surprising to witness that numerous Taiwan HEIs, facing the pressure from the Taiwan MOE and society, have required students to pass an English benchmark prior to graduation. In his article to mark the 25th anniversary of *Language testing*, Spolsky (2008) mentioned that the government and the general public substantially affect testing, both calling for using tests to hold education accountable and to implement educational reforms without devoting more resources. Although policymakers, educators, students, and the general public are confident about the impact of this policy, empirical studies have shown that this policy is merely a "placebo". The lack of washback stems from the complex nature of the washback phenomenon. In addition, although it plays a significant role in Taiwanese economy, English is not a skill that all Taiwanese students will use in their daily lives and workplaces. Other professional knowledge is, in fact, more critical than English, according to the interviewees in Chu's (2009) study. It is therefore untenable that students must attain a graduation benchmark in English, not on the expertise they pursue in their university studies. Finally, this policy sparks concerns about validity and ethics, as has been elaborated. In general, this policy does not achieve the goals established by policymakers, but it raises validity and ethical concerns. It is suggested that policymakers reconsider the English graduation benchmark.

Endnotes

- 1 Information on the GEPT can be found in some scholarly reviews (e.g., Roever & Pan, 2008; Shih, 2008b; Vongpumivitch, 2010) and on the LTTC's website (http://www.ltcc.ntu.edu.tw/ELTTC/gept_eng_main.htm).
- 2 The IELTS means the International English Language Testing System.
- 3 The TOEFL stands for the Test of English as a Foreign Language.
- 4 The TOEIC is the acronym of the Test of English for International Communication.

References

- Berry V & Lewkowicz J 2000. Exit-tests: Is there an alternative? *Hong Kong Journal of Applied Linguistics*, **5**(1):19-49.
- Brown HD 2004. *Language assessment: Principles and classroom practices*. New York: Pearson.
- Brown J D & Hudson T 1998. The alternatives in language assessment. *TESOL Quarterly*, **32**: 653-675.
- Chen T-H 2008. Impacts of compulsory standardized exams on college students' L2 learning motivation. MEd dissertation. Hsinchu: National Chiao Tung University.
- Chou C-I 2008. The impact of neo-liberalism on Taiwanese higher education. *International Perspectives on Education and Society*, **9**:297-311.
- Chu H-Y 2009. Stakes, needs and washback: An investigation of the English benchmark policy for graduation and EFL education at two universities of technology in Taiwan. PhD dissertation. Taipei: National Taiwan Normal University.
- Haddad WD & Demsky T 1995. *Education policy-planning process: An applied framework*. Paris: United Nations Educational, Scientific, and Cultural Organization.
- Language Training and Testing Center n.d.-a. *Advanced*. Retrieved on 18 September 2009 from http://www.ltcc.ntu.edu.tw/E_LTTC/gept_eng_ad.htm.
- Language Training and Testing Center n.d.-b. *High-intermediate*. Retrieved on 18 September 2009 from http://www.ltcc.ntu.edu.tw/E_LTTC/gept_eng_hi.htm.
- McNamara T 2008. The socio-political and power dimensions of tests. In NH Hornberger (Series Ed.) & E Shohamy & NH Hornberger (Vol. Eds), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (2nd ed., pp. 415-427). New York: Springer Science+ Business Media.
- McNamara T & Roever C 2006. *Language testing: The social dimension*. Malden, MA: Blackwell Publishing.
- Messing S 1996. Validity and washback in language testing. *Language Testing*, **13**:241-256.
- Ministry of Education 2005. 教育施政主軸行動方案 [Four-year plan for 2005-2008 in education]. Retrieved on 25 April 2005 from http://www.edu.tw/EDU_WEB/EDU_MGT/SECRETARY/EDU8354001/940309-01.doc.
- Ministry of Education n.d. 教育政策白皮書(初稿) [White Paper of educational policies (draft)]. Retrieved on 2 September 2011 from [mail2.ylc.edu.tw/~gces.kentxchang/pipi/white.doc](mailto:~gces.kentxchang/pipi/white.doc)
- National Yunlin University of Science and Technology 2008. 國立台中技術學院評鑑報告 [Evaluation report on National Taichung Institute of Technology]. Retrieved on 14 September 2009 from <http://www.twaea.org.tw/download/tech-school/台中.pdf>.
- Pan Y-C 2009. *The social impact of English certification exit requirement*. Paper presented at the Language Testing Research Colloquium, Denver. Retrieved on 26 September 2009 from http://www.cal.org/ltrc2009/schedule/ltrc2009_program.pdf.
- Roever C & Pan Y-C 2008. GEPT: General English Proficiency Test. *Language Testing*, **25**:403-408.
- Shih C-M 2007. A new washback model of students' learning. *Canadian Modern Language Review*, **64**:135-162.
- Shih C-M 2008a. Critical language testing: A case study of the General English Proficiency Test. *English Teaching and Learning*, **32**(3):1-34.
- Shih C-M 2008b. The General English Proficiency Test. *Language Assessment Quarterly*, **5**:63-76.
- Shih C-M 2010. The washback of the General English Proficiency Test on university policies: A Taiwan case study. *Language Assessment Quarterly*, **7**:234-254.

- Shohamy E 2001. *The power of tests: A critical perspective on the uses of language tests*. Essex: Longman.
- Spolsky B 1997. The ethics of gatekeeping tests: What have we learned in a hundred years? *Language Testing*, **14**:242-247.
- Spolsky B 2008. Language testing at 25: Maturity and responsibility? *Language Testing*, **25**:297-305.
- Taiwan Government Information Office 2004. *A brief introduction to Taiwan: The economy*. Retrieved on 17 August 2005 from http://www.gio.gov.tw/taiwan-website/5-gp/brief/info04_8.html.
- Taiwan Government Information Office 2008. *The Republic of China Yearbook 2008*. Retrieved on 15 September 2009 from <http://www.gio.gov.tw/taiwan-website/5-gp/yearbook/ch9.html>.
- Vongpumivitch V 2006. *An impact study of Taiwan's General English Proficiency Test (GEPT)*. Paper presented at the Language Testing Research Colloquium, Melbourne, Australia.
- Vongpumivitch V 2010. The General English Proficiency Test. In L Cheng & A Curtis (Eds), *English language assessment and the Chinese learner* (pp. 158-172). New York: Routledge.